An acoustic analysis of lion roars. II: Vocal tract characteristics

G. Ananthakrishnan [1], Robert Eklund [2,3,4], Gustav Peters [5] & Evans Mabiza [6]

[1] Centre for Speech Technology, KTH, Stockholm, Sweden
[2] Voice Provider, Stockholm, Sweden
[3] Department of Cognitive Neuroscience, Karolinska Institute, Stockholm, Sweden
[4] Department of Computer Science, Linköping University, Linköping, Sweden
[5] Forschungsinstitut Alexander Koenig, Bonn, Germany
[6] Antelope Park, Gweru, Zimbabwe

Abstract

*This paper makes the first attempt to perform an acoustic-to-articulatory inversion of a lion (*Panthera leo*) roar. The main problems that one encounters in attempting this, is the fact that little is known about the dimensions of the vocal tract, other than a general range of vocal tract lengths. Precious little is also known about the articulation strategies that are adopted by the lion while roaring. The approach used here is to iterate between possible values of vocal tract lengths and vocal tract configurations. Since there seems to be a distinct articulatory changes during the process of a roar, we find a smooth path that minimizes the error function between a recorded roar and the simulated roar using a variable length articulatory model.*

Introduction

The roar is a distinct mammalian vocalization made by only five species of *Felidae*. Researchers suggest that the ability to roar is made possible due to the specialized hyoid apparatus present in these mammals (Weissengruber et al., 2002). Acoustic-articulation modeling has been applied on several mammalian vocalizations in order to estimate the approximate vocal tract length of the animal producing the sound (Hauser, 1993; Taylor and Reby, 2010). The purpose has often been to correlate the estimated length of the vocal tract to the size of the animal to see if larger vocal tract lengths meant relative size dominance. The estimates were further correlated with the social behavior and mating roles of these vocalizations. Most of these methods applied the source-filter theory (Fant, 1970; Titze, 1994) to obtain inferences regarding the vocal tract characteristics. Here the properties of the larynx control the source signal characteristics, while the vocal tract configuration controls the filter characteristics. Since articulation data for mammals have not been very easy to obtain, most of these methods assume a uniform vocal tract for the mammals when they produce the sound and use the formant dispersion method (Titze, 1994; Fitch, 1997).

The lion roaring sequence usually consists of three different phases (Peters, 1978). The first phase is a series of low-intensity calls similar to 'mews'. The second phase, builds up to the climax with calls of increasing duration (shortening again towards the climax). Finally the sequence ends with a series of 'grunt' like sounds. In this study, we are interested in the second phase which is tonal in nature and has the maximum intensity in the entire sequence. Henceforth we only refer to the second phase by the word 'roar'.

Figure 1 shows the spectrogram of a prototypical roar of a female lion. It is clear that there is change in the formant structure also illustrated in *Figures* 2 and 3, showing the Spectral Envelopes varying over time and the average spectral slices for the two parts of a single roar respectively. This change in formant structure indicates that there is a corresponding change in the vocal tract dimensions during the process of producing the roar. Change in the quality of vocalizations have also been observed in other animals to where the vocalization includes protrusion of lips or jaw movement (e.g., Harris et al. 2006). Some species of fallow deer (*Dama dama*) are known to lower their larynx during the call (Vannoni et al., 2005).

Given this observation of changing formant structure during the roar, the uniform tube assumption can no longer be valid. One can suppose that that the filter (vocal tract) undergoes

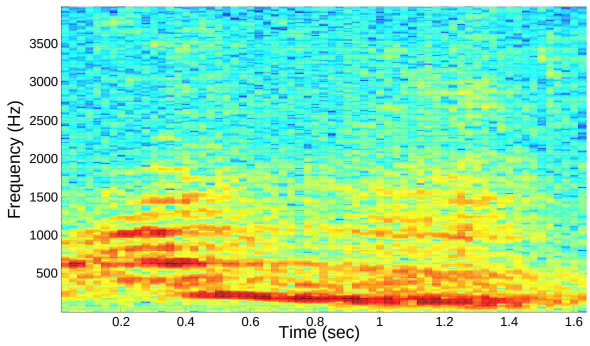


Figure 1: The spectrogram of a typical lion roar (in this case, a female lion's).

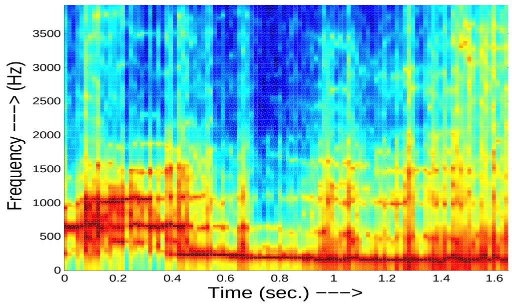


Figure 2: Illustration of the temporal changes in the formant structure, and therefore vocal tract configuration.

some change. However, one does not know what kind of change the vocal tract undergoes, whether it is the lowering of the larynx or changing of the vocal tract area function, or a combination of both.

Theory and Methods

The method proposed in this paper uses a Variable Linear Articulatory Model (VLAM) which allows the articulatory synthesizer developed by Maeda (1979) to be operated at different vocal tract lengths. Although this synthesizer has been designed for human-voices, the source-filter theory as shown previously by Taylor and Reby (2010) can be applied to other mammal vocalizations too. However, since the vocal tract area functions of a lion are largely unknown, we iterate over a range of values and select a configuration which best matches the spectral envelope of the recording of a lion roar. The several steps in the process are described below

1. The lion roar signal is segmented into overlapping windows, using the 'Hann' window function. Each window length is 30 ms in duration and successive windows are 5 ms apart.

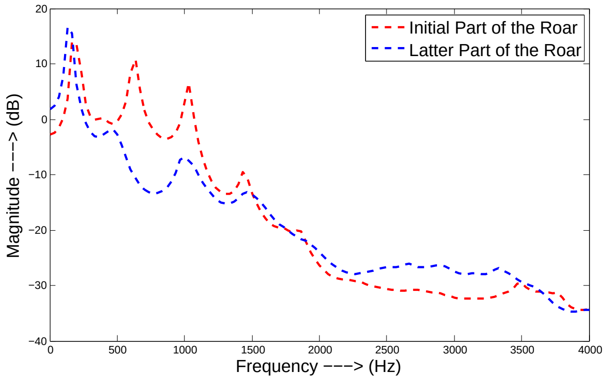


Figure 3: The Spectral Envelope, estimated using LPC analysis, from the beginning and the ending of one lion roar. This indicates that there is some change in the vocal tract configuration during the roar.

2. Linear Prediction Coefficients (LPC) were calculated for each window and then a Fast Fourier Transform (FFT) was applied, to the calculated transfer function so as to obtain the spectral envelope. The number of LPC parameters was set to 21, so as to obtain around 9 to 11 formant peaks within 4000 Hz. This was estimated based on the approximate dimensions of the Vocal-Tract Length (VTL) of a Lion, which is around 35 to 40 cm.

3. The spectral envelope for each window was converted to the decibel (dB) scale and normalized so as to limit the largest formant peak to 0 dB. We also subtracted the mean spectral slope from detected formants, so as to remove the effect of voicing in the estimates of the vocal tract shape.

4. We divided the vocal tract into three equal regions called the Jaw Section, Oral Section and the the Pharyngeal Section. The cross-sectional areas of the three sections were called JawSec, OralSec and PharSec respectively. We performed smoothing and linear interpolation on the three sections in order to approximate a 40 cylindrical tube model.

5. Using the VLAM simulations, we simulated the spectral transfer function, given different combinations of values for the four parameters VTL, JawSec, OralSec and PharSec. The spectral transfer function for each configuration was compared with the spectral envelope of the waveform for each time window to find the Euclidean distance between the two spectra.

6. Since several combinations of VTL and area functions can contribute to largely similar spectral characteristics Atal et al. (1978), we apply a smoothing function on the estimated vocal tract parameters. The movement being a muscular motion, a minimum jerk trajectory is the expected type of movement (at least for humans) Viviani and Terzuolo (1982). We thus apply a minimum jerk smoothing with multiple hypotheses Ananthakrishnan and Engwall (2011). The hypotheses are the 10 vocal tract configurations with minimum estimation error for each frame. These hypotheses are weighted by the inverse of the estimation error.

Data and Experiments

The data we used were recordings of lion roars made at two locations, namely, at the Antelope Park (Gweru, Zimbabwe), and Parken Zoo (Eskilstuna, Sweden). The equipment used at the Antelope Park was a DM50 electret stereo condenser shotgun microphone with a 150–15,000 Hz frequency range and a sensitivity of -40 dB. The estimated distance between the microphone and the lions varied from about four meters to ten meters, with the microphone pointing towards the general direction of a group of nine male lions (most of them born in 2006) in an open enclosure. Although there were other roars, we only considered the loudest roars which we assumed to be from the nine males mentioned above. The recordings at the Parken Zoo were made with two Canon HG-10 HD camcorders. One camera used the same microphone (DM50) as described above, while the other camera used an Audiotechnica AT813 cardoid-pattern, condenser mono microphone, with a frequency range of 30–20,000 Hz and a sensitivity of -44 dB. There where three lions, one male and two females. The male was 12 years old and weighed around 180 kilograms, while the females weighed around 165 kilograms was were around 14 years old. Further details of the data collected are mentioned in Eklund et al. (2011).

The waveforms were initially sampled at 44100 Hz, but were later sub-sampled to 8000 Hz to ensure compatibility with the VLAM model which estimated the vocal tract spectral transfer function in the frequency range of 0 to 4000 Hz. The waveforms were manually segmented to extract the second part of the roaring sequence, i.e. the tonal roar. We used a range of possible vocal tract lengths, ranging from 16 cm to 54 cm. The area functions for the three vocal tract

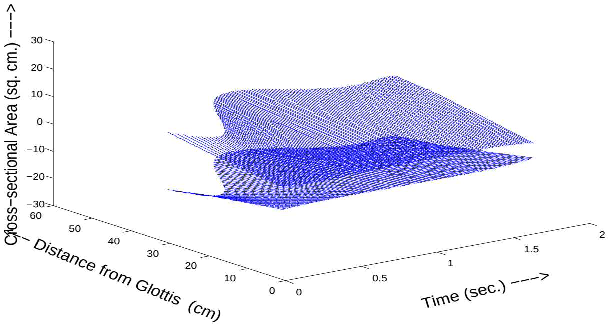
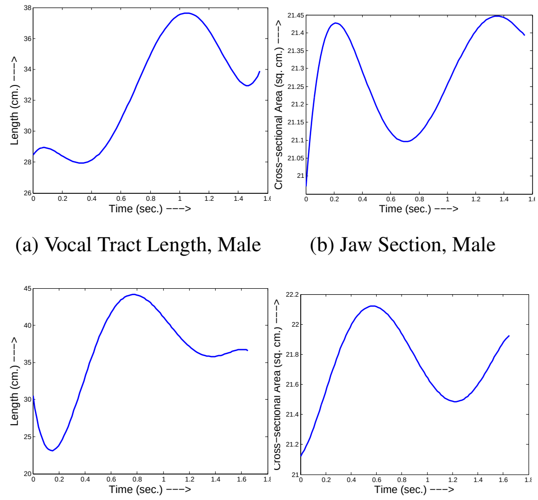


Figure 4: Illustration of how the vocal tract area function changes with respect to time during the course of a roar.



(a) Vocal Tract Length, Male (b) Jaw Section, Male

(c) Vocal Tract Length, Female (d) Jaw Section, Female

Figure 5: Illustration of how the vocal tract length and Jaw cross-sectional areas change with respect to time during the course of a roar.

sections were iterated between 8 to 24 sq. cm. These estimates were then compared with the videos sequences wherever available.

Results and Conclusions

Figures 4 and 5 indicate the estimated vocal tract shapes and VTLs over time for the female lion. This shows that vocal tract of a lion, approximates a frustum of a cone, rather than a uniform cylinder. The plots also indicate that, the roar involves a lengthening of vocal tract and then then a stabilization during the course of the roar. The range of variation is from 28 cm to 38 cm for the male lion and from 25 cm to 45 cm for the female lion. This may be effected by lowering the larynx, achieved during lifting up of the head. The female lion shows a larger variation in VTL during the course of the roar. The results

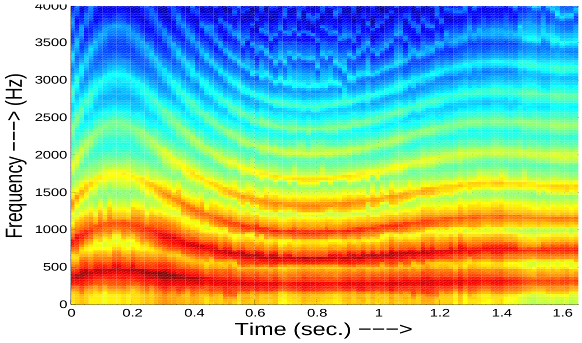


Figure 6: Illustration of the estimated spectral envelope of the lion roar.

also shows a slight decrease in the jaw area, especially for the female. In the videos that were recorded, the lions lifted their head up each time they roared. The jaw saw an increased opening followed by a reduction in the opening during the roars. The prediction from the estimates fit well with the observation about the VTL and the JawSec.

Dynamic analysis of animal vocalizations in order to extract the vocal tract characteristics is a very preliminary attempt in this paper. Some interesting observations have been uncovered in this study. The first being, the general shape of the vocal tract being more conical rather than cylindrical. Secondly, there seems to be a clear indication of larynx lowering, which is similar to the observations on fallow deer vocalizations (Vannoni et al., 2005). However, the female vocal tract is expected to be smaller than the male vocal tract given the differences in overall sizes. The mean VTL of the female lion's is estimated to be around 36 cm and is longer than the male lion's, estimated to be around 32 cm, which is rather unintuitive. Anatomical evidence for a male lion's vocal tract suggests a length of 38 cm Weissengruber et al. (2002). Estimating the mean VTL obscures the fact that the change in VTL for the female lion is also larger than the male lion's. This does not give any indication of what the static and normal lengths would be. Although some observations can be verified using video sequences, other observations need further data and analysis before make strong conclusions.

Future work would include analyzing physical, biological and ecological reasons for this type of motion during the roar, as well as other acoustic properties of the roar. Initial observations point to an increased roughness in the latter part of the roar, likely to be influenced by the voice source. This would also be an interesting investigation.

Acknowledgement

The authors would like to thank Jennie Westander and Conny Gärskog at Parken Zoo. Also thanks to Miriam Oldenburg for help with the "roarcordings" at Zimbabwe and Eskilstuna.

References

Ananthakrishnan G and Engwall O (2011). Mapping between acoustic and articulatory gestures. *Speech Communication*, 53(4):567–589.

Atal B S, Chang J J, Mathews M V and Tukey J W (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *The Journal of the Acoustical Society of America*, 63(5):1535–1555.

Eklund R, Peters G, Ananthakrishnan G and Mabiza E (2011). An acoustic analysis of lion roars. I: Data collection and spectrogram and waveform analyses. In *Proc. Fonetik 2011*, this volume. Stockholm, Sweden.

Fant G (1970). *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. Mouton De Gruyter.

Fitch W T (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J Acoust Soc Am*, 102:1213–1222.

Harris T, Fitch W, Goldstein L and Fashing P (2006). Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. *Ethology*, 112(9):911–920.

Hauser M D (1993). The evolution of nonhuman primate vocalizations: effects of phylogeny, body weight and social context. *American Nature*, 142:528542.

Maeda S (1979). An articulatory model of the tongue based on a statistical analysis. *J of Acous Soc Am*, 65:S22.

Peters G (1978). *Vergleichende Untersuchung zur Lautgebung einiger Feliden (Mammalia, Felidae)*. Zoologische Staatssammlung M
"unchen.

Taylor A M and Reby D (2010). The contribution of sourcefilter theory to mammal vocal communication research. *Journal of Zoology*, 208:221–236.

Titze I R (1994). *Principles of vocal production*. Englewood Cliffs: Prentice-Hall.

Vannoni E, Torriani M and McElligott A G (2005). Acoustic signaling in cervids: a methodological approach for measuring vocal communication in fallow deer. *Cognition, Brain, Behavior*, IX:551–565.

Viviani P and Terzuolo C (1982). Trajectory determines movement dynamics. *Neuroscience*, 7(2):431–437.

Weissengruber G, Forstenpointner G, Peters G, Kübber-Heiss A and Fitch W T (2002). Hyoid apparatus and pharynx in the lion (*Panthera leo*), jaguar (*Panthera onca*), tiger (*Panthera tigris*), cheetah (*Acinonyx jubatus*) and domestic cat (*Felis silvestris f. catus*). *Journal of anatomy*, 201(3):195–209.

KTH Computer Science
and Communication

Speech, Music and Hearing

Quarterly Progress and Status Report

TMH-QPSR, Volume 51, 2011

Proceedings from

Fonetik 2011

June 8 – June 10, 2011

Organized by
Department of Speech, Music and Hearing,
and Centre for Speech Technology, CTT,
KTH, Stockholm